計算の複雑さの構造的解析法の研究とその応用
－計算論的学習理論に関する研究成果－

渡辺　治

東京工業大学　情報理工学研究科

計算工学専攻　助教授

1．はじめに

コンピュータにいろいろな処理をさせる場合、その処理が難しい問題もあれば容易な問題もある。しかし、この一見当たり前に思えることでも、では何故、ある問題の計算が難しく、また一方で（似たような）別の問題の計算がやさしいのか、といった疑問に正確に答えるのは非常に難しい。しかし、問題の難しさの本質が明確にならない限り、よい解決方法（アルゴリズム）を科学的に構成することは不可能に近い。そこで、「計算の難しさの構造的解析」と呼ばれている研究分野では、問題の難しさの本質を探る方法を開発し、それによって難しさの本質を突き止める研究をしている。本研究では、構造的解析の手法を計算論的学習理論に応用し、代表的な帰納学習の性能と問題点を調べた。

計算論的学習理論は、従来、経験的あるいは実験科学的に研究されてきた学習（主に機械学習）に関する研究を、計算手法（アルゴリズム）という観点から、より数学的に解明しようという研究分野である。この分野では、いろいろな学習過程をコンピュータの学習問題としてモデル化し議論している。本研究では、代表的なモデルの１つである「質問による学習」（query learning）について、その難しさや学習アルゴリズムについて考察した。

2．本研究で考える「学習」とは

一般には「学習」というと、人間が行なうような万能型の学習がイメージされる。しかし、ここで考える学習は、もっとずっと単純で、問題領域を非常に限定した中での学習である。つまり、学習対象がある特定の型に定まっていて、その中で（コンピュータなどに）与えられた対象が何であるかを同定させることを考えている。

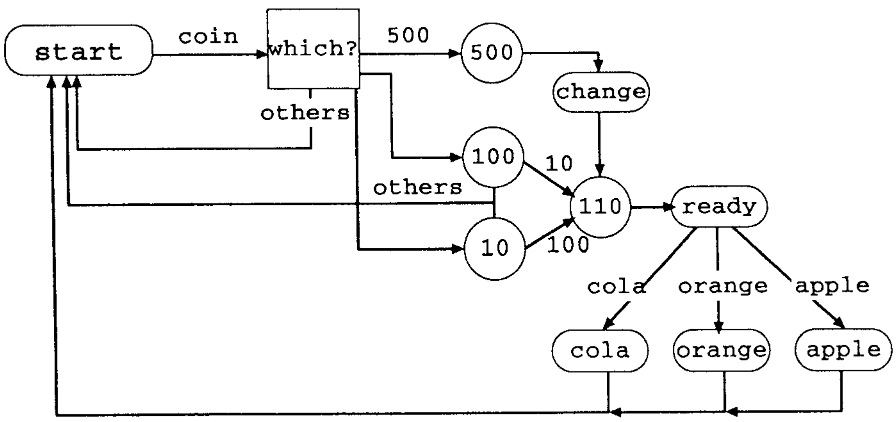

図１：単純なジュースの自動販売機の構造

注）ここでは図を簡単にするために、実際の自動販売機を非常に単純化したものを考えている

一例としてジュースの自動販売機の学習を考えてみよう。これはジュースの自動販売機の内部構造を、図１のような流れ図として表す問題である。もちろん、与えられたジュースの自動販売機の中身を見れば、それから図１のような流れ図を作るのは、そう難しいことではない。そうした完全な内部情報がない状態で、限られた（間接的な）部分情報から、流れ図を構成するのが我々の学習問題である。

たとえば、対象となるジュースの自動販売機のことをよく知っている人がいたとしよう。その人（これを**教師**という）に対して、「100円を入れたらどうなりますか？」といった質問を繰り返し、その答えを総合して流れ図を作成する。これが本研究の対象である**質問による学習**（以下では単に**質問学習**と呼ぶ）である。その他にも、実際にジュースの自動販売機に対して、いろいろな操作をランダムに行ない、その結果の観察から内部構造を推定する学習もある。これは**ＰＡＣ学習**（Probably Approximately Correct Learning）と呼ばれている。

このように、「学習」とはいっても、一般にイメージされる学習とは異なり、かなり限定された形の学習である。しかし、こうした限定された学習問題でも、コンピュータの利用技術の進歩に、いろいろな形でかかわってくることができる。

その一例として、コンピュータを利用した要求仕様書の作成方法を考えてみよう。上記の流れ図は、一般には**オートマトン**（有限状態機械）と呼ばれている。実際の応用場面では、作成したい機械、ソフトウェア、あるいは通信システムなどが、このオートマトンで記述できる場合がしばしばある。しかし、その機械の注文者に、機械の内部状態をオートマトンの形で形式的に記述するよう要請しても、難しい場合が多い。そこで、もしオートマンを質問学習により構成するプログラムがあれば、機械の注文者を教師に見立て、注文者にいろいろと質問を繰り返しながら、所望の機械の使用を要領良くオートマトンにまとめることも可能だろう。つまり、コンピュータの助けを借りながら、その機械の要求仕様を厳密に作成することができるのである。しかしそのためには、より少ない質問回数で、しかも効率よく、オートマトンを作成する計算手法（アルゴリズム）が重要になってくる。それが、以下に述べる質問学習の効率に関する研究の主な動機である。

３．オートマトンの質問学習の効率について

Angluin〔1〕は、オートマトンを多項式時間以内に同定する質問学習のアルゴリズムを発見した。この学習アルゴリズムでは、所属質問と同値質問が用いられる。**所属質問**とは、入力例に対してオートマトンがどのように反応するかを聞く質問である。たとえば、先の自動販売機の例で言えば、「500円を入れるとどうなるか？」というような質問である。一方、**同値質問**とは、オートマトンの状態遷移図を仮説として提示して、それが所望のオートマトンであるかを聞く質問のこと。たとえば、自動販売機の流れ図を示し、これが正しい内部構造であるかを聞くよな質問である。（先に述べたような応用では、同値質問をするのが難しい場合が考えられる。そのような場合には、直接教師に同値質問をするのではなく、いくつかの入力例により実験を行なうことにより、仮説のオートマトンの整合性をチェックする手法が妥当だろう。）

Angluin〔1〕の提案したアルゴリズムでは、状態数ｎのオートマトンを同定するのに、$O(n)$回の同値質問と、$O(n^2)$回の所属質問が必要になる（なお、計算時間は$O(n^3)$程度）。本研

究では、この質問回数が改善できないかを調べた。

以下では次のような記法を用いる：オートマトンの学習アルゴリズムAに対し、

$\#mem - query_A(n)$ = 状態数 n のオートマトンを学習するために必要な所属質問の数。

$\#equ - query_A(n)$ = 状態数 n のオートマトンを学習するために必要な同値質問の数。

まず、同値質問が本質的であるかについて調べた。同値質問は、所属質問に比べてはるかに実現にコストがかかる。したがって、同値質問がなるべく少ない方がよい。しかし比較的簡単な議論で、同値質問は必要であることが証明できる。我々は、学習アルゴリズムをランダム化（randomized）し、多少誤ることを許したとしても、同値質問はやはり本質的であることを証明した。

定理１． 任意のオートマトンの学習アルゴリズムAを考える。ただし、Aでは、ランダム化を許し、適当な定数δ以下の誤り率を許すことにする。このとき、もしAが所属質問だけしかしないとすると、その回数に関して次の下限が成り立つ。

$$\#mem - query_A(n) \geq \delta 2^n - 1.$$

このように、オートマトンの学習には、同値質問が本質的なのであるが、実は、所属質問も同様に本質的であることが、Angluin〔2〕によって示されている。このことから、所属質問と同値質問の間にはトレード・オフの関係があるものと予想される。本研究では、これを以下に示す２つの定理により実証した（なお、最初の定理の証明では構造的解析において使われている、アドバーサリー法を用いた）。

定理２． Aを、オートマトンの学習アルゴリズムとする。またｆ(ｎ)＜ｎを任意の非定数関数とする。このとき、ある定数c_1, c_2に対し、次のいずれかが成り立つ：

$$\#equ - query_A(n) \geq \frac{n}{f(n)}, \quad or$$
$$\#mem - query_A(n) \geq 2^{c_1 f(n)} - \frac{c_2 n}{f(n)}.$$

定理３． 任意のｆ(ｎ)＜ｎに対し、次の評価式を満たすようなオートマトンの学習アルゴリズムA_fを構成できる：

$$\#equ - query_{Af}(n) \leq \frac{n}{f(n)}, \quad and$$
$$\#mem - query_{Af}(n) \leq 2^{f(n)} q_f(n).$$

ただし q_f は適当な多項式。

たとえば、定理2より、まともな計算時間（多項式時間）内にオートマトンを同定するためには、どうしても$n/\log n$回の同値質問は必要であることがわかる。一方、$n/\log n$回の同値質問で十分多項式時間以内に同定できることが、定理3によって示されている。

なお、ここまでの結果の詳細は論文〔4〕を参照されたい。

さらに我々は、並列計算でオートマトンを同定する手法について考察した。実は、定理3の手法は並列計算に比較的むいており、それをうまく改良すると約$O(n/\log n)$時間の並列学習アルゴリズムが作れる。一方、どのような並列アルゴリズムを考えても、同値質問の回数$n/\log n$だけは、どうしても時間がかかってしまうことも証明できた。すなわち、$O(n/\log n)$時間の並列学習アルゴリズムは最適なアルゴリズムなのである。正確には、次の定理が得られた（詳しくは〔5〕参照）。

定理4． オートマトンの学習をする$O(n/\log n)$時間のPRAMアルゴリズムが構成できる。しかも、オートマトンの学習をするPRAMアルゴリズムで、計算時間が$O(n/\log n)$のものは存在しない。

4．質問学習の難しさについて

いままで述べてきたように、オートマトンの質問学習には、かなり効率のよいアルゴリズムが存在する。ところが、実際の応用面では、オートマトンよりもう少し複雑なメカニズムの同定できれば都合のよい場合が、しばしば生じる。たとえば、文脈自由文法などが同定できるようになると、自然言語処理への応用の可能性も広がる。そうした、より一般的なメカニズムの学習は、どの程度可能なのだろうか？

こうした研究のため、渡辺〔6、7〕は、構造的解析の手法を利用して、質問学習可能性を調べるための枠組を提案した。その枠組にそった研究の結果が、以下に述べる2つの定理である。なお、以下では、説明を単純にするために、多少簡単化している部分もある。詳しくは論文〔8〕を参照されたい。

定理5． 文脈自由文法を質問学習アルゴリズムは、（教師の助けを借りても）高々$P(NP^{(\ )})$程度の計算能力しかもたない。さらに、性質Rを持つ文脈自由文法を質問学習する場合には、ちょうど$P(NP^{(\ )})$計算能力を持つことができる（なお、性質Rを簡単にいうと「繰り返し文を内包している」という性質）。

すなわち、ある種の文脈自由文法で、その学習問題の難しさが$P(NP^{(\ )})$を越えたとすると、この定理より、その種の文脈自由文法を、多項式時間以内に同定するのは不可能になる。一方、性質Rを持つ文脈自由文法の学習不可能性が証明できたとすると、難しさが$P(NP^{(\ )})$を越える学習問題を示すことができる。

このように、文脈自由文法を同定する学習問題を考えると、その計算の複雑さが、多項式時間階層のどこに位置するかが、文脈自由文法の学習可能性に結び付いてくる。では、その問題が実

際に多項式時間階層のどこに位置するかを考えてみると、次のような定理が示せる。

定理6． 文脈自由文法を同定する学習問題の難しさは、高々$\mathrm{P}(\mathrm{NP}(\mathrm{NP}^{(\ )}))$ 程度である。

しかし、文脈自由文法の学習問題が、難しさが$\mathrm{P}(\mathrm{NP}^{(\ )})$ を越えるか否かは、まだ未解決である。

5．おわりに

計算の複雑さの構造的解析の手法を用いて、質問による学習の効率の評価、ならびに学習可能性の検討を行なった。今後の課題としては、まず第一に、文脈自由文法の質問学習の可能性を、さらに深く追求することが重要である。ただし、今までの研究からの経験では、文脈自由文法がすべて質問学習可能であるかは、かなり疑わしい。そこで、どの程度の文脈自由文法が学習可能かが重要になってくるだろう。とくに、平均的な学習可能性の解析が興味深い。その他の課題としては、ＰＡＣ学習などのような他の学習モデルと、質問学習を融合した形の学習モデルの解析もおもしろいだろう。

最後に、本研究の遂行に多大なご援助を頂きました㈶高柳記念電子科学技術振興財団ならびに財団の関係者の皆様に深くお礼申し上げます。

参考文献

〔1〕D. Angluin, Learning regular sets from queries and counterexamples, Information and Computation, Vol.75, pp.87-106 (1987).

〔2〕D. Angluin, Negative results for equivalence queries, Machine Learning, Vol. 5, pp. 121-150 (1990).

〔3〕J. Balcázar, J. Díaz, R. Gavaldà, and O. Watanabe, “A note on the query complexity of learning DFA”, in Proc. Third Workshop on Algorithmic Learning Theory, Japanese Society for Artificial Intelligence, pp.53-62 (1992).

〔4〕J. Balcázar, J. Díaz, R. Gavaldà, and O.Watanabe, The query complexity of learning DFA, New Generation Computing, Vol. 12, pp.-337-358 (1994).

〔5〕J. Balcázar, J. Díaz, R. Gavaldà, and O. Watanabe An optimal parallel algorithm for learning DFA, in Proc. 7th ACM Conference on Computational Learning Theory, ACM, New York, 208-217 (1994).

〔6〕O. Waranabe, A formal study of learning via queries, in Proc. 17th International Colloquium on Automata, Languages and Programming, Lecture Notes in Computer Science 443, Springer-Verlag, pp. 137-152 (1990).

〔7〕O. Watanabe A framework for polynomial time query learnability, Mathematical Systems Theory, Vol. 27, 211-229 (1994).

〔8〕R. Gavaldà and O. Watanabe, Structural analysis of polynomial time query learnability, Mathematical Systems Theory, Vol. 27, 231-256 (1994).